Rank-preserving geometric means of positive semi-definite matrices

Silvere Bonnabel* Anne Collard Rodolphe Sepulchre †

March 29, 2012

Abstract

The generalization of the geometric mean of positive scalars to positive definite matrices has attracted considerable attention since the seminal work of Ando. The paper generalizes this framework of matrix means by proposing the definition of a rank-preserving mean for two or an arbitrary number of positive semi-definite matrices of fixed rank. The proposed mean is shown to be geometric in that it satisfies all the expected properties of a rank-preserving geometric mean. The work is motivated by operations on low-rank approximations of positive definite matrices in high-dimensional spaces.

1 Introduction

Positive definite matrices have become fundamental computational objects in many areas of engineering and applied mathematics. They appear as covariance matrices in statistics, as variables in convex and semidefinite programming, as unknowns of important matrix (in)equalities in systems and control theory, as kernels in machine learning, and as diffusion tensors in medical imaging, to cite a few. These applications have motivated the development of a difference and differential calculus over positive definite matrices. As a most basic operation, this calculus requires the proper definition of a mean. In particular, much research has been devoted to the matrix generalization of the geometric mean $\sqrt{ab}$ of two positive numbers a and b (see for instance Chapter 4 in [1] for an expository and insightful treatment of the subject). The further extension of a geometric mean from two to an arbitrary number of positive definite matrices is an active current research area [2, 3, 4]. It has been increasingly recognized that from a theoretical point of view [5] as well as in numerous applications [6, 7, 8, 9, 2, 10, 3, 11, 12, 13], matrix geometric means are to be preferred to their arithmetic counterparts for developing a calculus in the cone of positive definite matrices.

The fundamental and axiomatic approach of Ando [2] (see also [3]) reserves the adjective "geometric" to a definition of mean that enjoys all the following properties:

(P1) Consistency with scalars: if A and B commute, then $M(A,B)=(AB)^{1/2}$.

(P2) Joint homogeneity

$$M(\alpha A,\beta B)=(\alpha\beta)^{1/2}M(A,B).$$

*S. Bonnabel is with Centre de Robotique, Mathématiques et Systèmes, Mines ParisTech, Boulevard Saint-Michel 60, 75272 Paris, France. silvere.bonnabel@mines-paristech.fr

†Anne Collard and Rodolphe Sepulchre are with Departement of Electrical Engineering and Computer Science, University of Liège, 4000 Liège, Belgium. {Anne.Collard,r.sepulchre}@ulg.ac.be

(P3) Permutation invariance $M(A, B) = M(B, A)$.

(P4) Monotonicity. If $A \leq A_0$ (i.e. $(A_0 - A)$ is a positive matrix) and $B \leq B_0$, the means are comparable and verify $M(A, B) \leq M(A_0, B_0)$.

(P5) Continuity from above. If $\{A_n\}$ and $\{B_n\}$ are monotonic decreasing sequence (in the Lowner matrix ordering) converging to A, B then $\lim(A_n \circ B_n) = M(A, B)$.

(P6) Congruence invariance. For any $G \in \mathrm{Gl(n)}$ we have $M(GAG^T, GBG^T) = GM(A, B)G^T$.

(P7) Self-duality $M(A, B)^{-1} = M(A^{-1}, B^{-1})$.

The present paper seeks to extend geometric means defined on the open cone $\mathrm{P_n}$ to the the set of positive semi-definite matrices of fixed rank p, denoted by $\mathrm{S^+(p, n)}$, which lies on the boundary of $\mathrm{P_n}$. Our motivation is primarily computational: with the growing use of low-rank approximations of matrices as a way to retain tractability in large-scale applications, there is a need to extend the calculus of positive definite matrices to their low-rank counterparts. The classical approach in the literature is to extend the definition of a mean from the interior of the cone to the boundary of the cone by a continuity argument. As a consequence, this topic has not received much attention. This approach has however serious limitations from a computational viewpoint because it is not rank-preserving. For instance Ando's geometric mean of two semi-definite matrices having rank $p < n/2$ is almost surely null with this definition.

We depart from this approach by viewing a rank p positive semi-definite matrix as the projection of a positive definite matrix in a p-dimensional subspace. Our proposed mean lies in the mean subspace, a well-defined and classical concept. The proposed mean is rank-preserving, and it possesses all the properties listed above, modulo a few adaptations imposed by a rank-preserving concept: (P1) is impossible unless the rank of AB is equal to the rank of A and B. Indeed, as the mean must preserve the rank, it can not be equal to $(AB)^{1/2}$ if the latter condition is not satisfied. Also (P6) will be shown to be impossible to retain when the rank is preserved. Indeed it is this property that causes Ando's geometric mean of two matrices of sufficiently small rank to be almost surely null. In (P7) inversion must obviously be replaced with pseudo-inversion. Letting $A \circ B$ denote the desired mean in $\mathrm{S^+(p, n)}$, we suggest to replace those three properties with:

(P1') Consistency with scalars: if A, B commute and AB has rank p, $A \circ B = (AB)^{1/2}$.

(P6') Invariance to scalings and rotations. For $(\mu, P) \in \mathbb{R}_+^* \times \mathrm{O(n)}$ we have $(\mu P^T A \mu P) \circ (\mu P^T B \mu P) = \mu P^T (A \circ B) \mu P$.

(P7') Self-duality $(A \circ B)^\dagger = A^\dagger \circ B^\dagger$, where $\dagger$ is the pseudo-inversion.

In the recent work [14], we used a Riemannian framework to introduce a geometric mean of two matrices in $\mathrm{S^+(p, n)}$ that was shown to satisfy those properties. The present paper further develops the concept by providing an inutitive characterization and a closed formula for its calculation. Furthermore, we show that the concept extends to the definition of a geometric mean for an arbitrary number of matrices, thereby providing the low-rank counterpart of recent work on positive definite matrices [2, 3, 4].

The structure of the paper is as follows. Section 2 and 3 are mainly expository. In Section 2, we review the theory of Ando in the cone of positive definite matrices and we illustrate the shortcomings of the continuity argument for a rank-preserving mean to be defined on the boundary of the cone. In Section 3, we review the Riemannian interpretation of Ando's mean of two matrices A and B as the midpoint of the geodesic joining A and B for the affine invariant metric of the cone and introduce the Riemannian concept of Karcher

mean. Section 4 develops the proposed geometric mean for an arbitrary number of matrices in the set $\mathrm{S}^+(\mathrm{p},\mathrm{n})$. The geometric properties of this mean are characterized in Section 5. Finally, Section 6 illustrates the relevance of a rank-preserving mean in the context of filtering. Preliminary results can be found in [15].

1.1 Notation

- $\mathrm{P_n}$ is the set of symmetric positive definite $n \times n$ matrices.
- $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ is the set of symmetric positive semidefinite $n \times n$ matrices of rank $p \leq n$.
- $\mathrm{Sym(n)}$ is the vector space of symmetric $n \times n$ matrices.
- $\mathrm{St(p,n)} = \mathrm{O(n)/O(n-p)}$ is the Stiefel manifold; i.e., the set of $n \times p$ matrices with orthonormal columns: $U^T U = I_p$.
- $\mathrm{Gr(p,n)}$ is the Grassman manifold, that is, the set of p-dimensional subspaces of $\mathbb{R}^n$. It can be represented by the equivalence classes $\mathrm{St(p,n)/O(p)}$.
- $\mathrm{span}(A)$ is the subspace of $\mathbb{R}^n$ spanned by the columns of $A \in \mathbb{R}^{n\times n}$.
- $T_X\mathcal{M}$ is the tangent space to the manifold $\mathcal{M}$ at X.

2 Analysis pass: Ando's approach

2.1 Mean of two matrices A_1 and A_2

For positive scalars, the homogeneity property (P2) implies $M(a_1, a_2) = a_1 M(1, a_2/a_1) = a_1 f(a_2/a_1)$ with f a monotone increasing continuous function. In a non-commutative matrix setting, one can write

$$M(A_1, A_2) = A_1^{1/2} f(A_1^{-1/2} A_2 A_1^{-1/2}) A_1^{1/2}\,. \tag{1}$$

with f a matrix monotone increasing function. Several geometric means can be defined in this way (see, e.g., [10]). The well-established geometric mean of two full-rank matrices popularized by Ando [16, 17, 2] corresponds to the case $f(X) = X^{1/2}$, generalizing the scalar geometric mean. It writes

$$A_1 \# A_2 = A_1^{1/2} (A_1^{-1/2} A_2 A_1^{-1/2})^{1/2} A_1^{1/2}. \tag{2}$$

It satisfies all the propositions (P1-P7) listed above. There are many equivalent definitions of the Ando geometric mean in the literature.

A geometric mean satisfying (1) is defined for positive definite matrices, that is, for elements in the open cone of positive definite matrices. Rank-deficient matrices lie on the closure of the cone. As a consequence, the natural idea to extend a geometric mean on the boundary is to use a continuity argument. The resulting mean satisfies all the properties above (except for (P7) that must be formulated using pseudo-inversion), but it is not rank-preserving. Indeed, let $A_1 = \mathrm{diag}(4, \epsilon^2)$ and $A_2 = \mathrm{diag}(\epsilon^2, 1)$ where the term $\epsilon \ll 1$. These two matrices belong to P_2, and their geometric mean is $A_1 \# A_2 = \mathrm{diag}(2\epsilon,\ \epsilon)$. In the limit (rank-deficient) situation $\epsilon \to 0$, the mean becomes the null matrix $\mathrm{diag}(0, 0)$. The following proposition shows that this example is not pathological.

Proposition 1. *If (P6) is satisfied, the geometric mean of two matrices of* $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ *is almost surely null if* $p < n/2$.

Proof. In [2], it is proved (Theorem 3.3) that (P6) implies that the range of the geometric mean of A_1 and A_2 is the intersection of the subspaces $\mathrm{span}(A_1)$ and $\mathrm{span}(A_2)$ (this can be proved letting a sequence of matrices of $\mathrm{Gl}(n)$ converge to the orthoprojector on Ker A_1). Since the intersection of two random subspaces of dimension p is almost surely empty as long as $n - p > p$, the range of $\mathrm{span}(A_1) \cap \mathrm{span}(A_2)$ is almost surely the null space, which proves the claim. $\square$

A rank-preserving mean thus requires a different approach. We seek to retain most of the properties (P1-P7), but we will see that three of them must be relaxed to define a rank-preserving mean. The major relaxation consists in choosing a smaller invariance group in (P6), replacing the general linear group Gl(n) with the smaller but meaningful group of scaling and rotations $\mathbb{R}_+^* \times O(n)$.

2.2 Mean of an arbitrary number of matrices $A_1, \cdots, A_N$

A geometric mean of an arbitrary number of matrices, that extends the geometric mean of two matrices (2) is not very well-established. Indeed the definitions based on equations (1) for instance, are not easily generalized. Several possible definitions have appeared in the literature and we shall not review all of them. In any case, it seems natural to reserve the adjective geometric, to a mean that satisfies the following properties (PP1-PP7). They are a natural extension of (P1-P7) to the case of three matrices, and the extension to an arbitrary number of matrices is straightforward. (PP1) if A,B,C commute $M(A, B, C) = (ABC)^{1/3}$. (PP2) $M(\alpha A, \beta B, \gamma C) = (\alpha\beta\gamma)^{1/3} M(A, B, C)$. (PP3) $M(A, B, C) = M(\pi(A, B, C))$ for any permutation π. (PP4) The map $(A, B, C) \mapsto M(A, B, C)$ is monotone. (PP5) If $\{A_n\}$, $\{B_n\}$, $\{C_n\}$ are monotonic decreasing sequences converging to A, B, C then $\lim M(A_n, B_n, C_n) = M(A, B, C)$. (PP6) For any $G \in$ Gl(n) we have $M(GAG^T, GBG^T, GCG^T) = GM(A, B, C)G^T$. (PP7) $M(A, B, C)^{-1} = M(A^{-1}, B^{-1}, C^{-1})$.

This axiomatic approach has been proposed in [2], and the authors have defined a mean we shall denote $\mathrm{alm}(A_1, \cdots, A_n)$, adopting the notation of [4]. This mean is defined as the common limit of a converging sequence of matrices, and it was proved to preserve properties (P1-P7) as well as their extension (PP1-PP7) to three or more matrices.

3 Geometric pass: geometric mean as a Riemannian mean

Ando's mean (2) has the alternative Riemannian interpretation of the midpoint of a geodesic connecting the matrices A and B. This connection appears for instance in [17]. Because this Riemannian interpretation is at the root of our proposed rank-preserving mean, it is reviewed in this section.

3.1 Riemannian mean and Karcher mean on a Riemannian manifold

The arithmetic mean of N positive numbers in $\mathbb{R}_+^*$ is defined as $M(x_1, \cdots, x_N) = \frac{1}{n}\sum_1^n x_i$. It has the variational property of being the unique minimizer of the sum of squared distances $M(x_1, \cdots, x_N) = \mathrm{argmin}_x \sum_i d(x, x_i)^2$ where d is the Euclidean distance in $\mathbb{R}$. In the same way, the geometric mean of n positive scalars minimizes the same sum if one works with the hyperbolic distance $d(x, y) = \mid \log x - \log y \mid$.

This variational approach allows to define candidate means of an arbitrary number of matrices on any connected Riemannian manifold $\mathcal{M}$. Such manifolds carry the structure of a metric space whose distance function is the arclength of a minimizing path between

two points. Indeed the mean of $x_1, \cdots x_N$ on $\mathcal{M}$, can be defined as the minimizer of the sum of squares $\sum_i d(x, x_i)^2$ where d is the geodesic distance on $\mathcal{M}$. Such a mean is known as the Riemannian barycenter, of Karcher or Fréchet mean. When only two points are involved, the Karcher mean is the midpoint of the minimizing geodesic connecting those two points and it is usually called the Riemannian mean. The main advantage of the Karcher mean is to readily extend any mean that can be defined as a geodesic midpoint, to an arbitrary number of points. Unfortunately the mean can rarely be given in closed form, and is typically computed by an optimization algorithm on the manifold (see e.g. [18] for more information on this branch of optimization). In [19] it has been shown that the Karcher mean is uniquely defined on manifolds with non-positive curvatures. On arbitrary manifolds with upper bounded sectional curvature, the Karcher mean exists and is unique in geodesic balls with sufficiently small radius [20].

3.2 Ando's mean as a Riemannian mean in the cone $\mathrm{P_n}$

Any positive definite matrix admits the factorization $X = YY^T$, $Y \in Gl(n)$, and the factorization is invariant by rotation $Y \mapsto YO$. As a consequence, the cone of positive definite matrices has a homogeneous representation $Gl(n)/O(p)$. The space is reductive and thus admits a Gl(n)-invariant metric called the natural metric of the cone of positive definite matrices [5]. If X_1, X_2 are two tangent vectors at $A \in \mathrm{P_n}$, the metric is given by $g_A^{P_n}(D_1, D_2) = \mathrm{Tr}\left(D_1 A^{-1} D_2 A^{-1}\right)$. With this definition, a geodesic (path of shortest length) at arbitrary $A \in \mathrm{P_n}$ is [9, 13]: $\gamma_A(tX) = A^{1/2} \exp(tA^{-1/2} X A^{-1/2}) A^{1/2}, \quad t > 0$, and the corresponding geodesic distance is

$$\begin{aligned} d_{P_n}(A, B) = d(A^{-1/2} B A^{-1/2}, I) &= \|\log(A^{-1/2} B A^{-1/2})\|_F, \\ &= \sqrt{\sum_k \log^2(\lambda_k)}, \end{aligned}$$

where λ_k are the generalized eigenvalues of the pencil $A - \lambda B$, i.e., the roots of $\det(AB^{-1} - \lambda I)$. The distance is invariant with respect to action by congruence of Gl(n) and matrix inversion.

The geodesic characterization provides a closed-form expression of the Riemannian mean of two matrices $A, B \in \mathrm{P_n}$. The geodesic $A(t)$ linking A and B is

$$A(t) = \exp_A^{P_n}(tX) = A^{1/2} \exp(t \log(A^{-1/2} B A^{-1/2})) A^{1/2},$$

where $A^{-1/2} X A^{-1/2} = \log(A^{-1/2} B A^{-1/2}) \in \mathrm{Sym(n)}$. The midpoint is obtained for $t = 1/2$: $M(A, B) = A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}$ and it corresponds to the definition (2).

3.3 Mean of positive definite matrices and Karcher mean in the cone $\mathrm{P_n}$

For $A_1, ..A_N \in \mathrm{P_n}$ viewed as a Riemannian manifold endowed with the natural metric, the Karcher mean is defined as a minimizer of $X \mapsto \sum_1^N d(X, A_i)^2$, i.e. a least squares solution that we shall denote $\mathrm{ls}(A_1, \cdots, A_n)$ as in [4]. The manifold $\mathrm{P_n}$ endowed with the natural metric is complete, simply connected, and has everywhere a negative sectional curvature. As a consequence, the Karcher mean is uniquely defined. It has been proposed by [8] as a natural candidate for generalizing the Ando mean to N matrices, and studied by [4]. It can mainly be calculated via a simple Newton method on $\mathrm{P_n}$, or by a stochastic gradient algorithm [21]. However, finding a closed-form expression of the Karcher mean of three or more matrices of $\mathrm{P_n}$ remains an open question. Several recent papers adress the issue of approximating the Karcher mean via simple algorithms [2, 3].

3.4 Mean of projectors and Karcher mean in the Grassman manifold

The Riemmanian approach to the definition of means provides a natural definition for the mean of p-dimensional projectors in $\mathbb{R}^n$, which forms a subset of $\mathrm{S}^+(\mathrm{p},\mathrm{n})$:

$$\{P \in \mathbb{R}^{n\times n} /\ P^T = P,\ P^2 = P,\ \mathrm{Tr}\,(P) = p\}, \tag{3}$$

This set is in bijection with the Grassman manifold of p-dimensional subspaces Gr(p,n) (e.g. [18]). This manifold can be endowed with a natural Riemannian structure. The squared distance between two subspaces is merely the sum of the squares of the principal angles between those two p-planes. The Riemannian mean of two subspaces is uniquely defined as soon as all the principal angles between those subspaces are stricly smaller than $\pi/2$. In other words, the injectivity radius at any point, i.e. roughly speaking the distance at which the geodesics cease to be minimizing, is equal to $\pi/2$ on this manifold. The Karcher mean of N subspaces $S_1, \cdots S_N$ of $\mathrm{Gr}(p,n)$ is defined as the least squares solution that minimizes $X \mapsto \sum_1^N d_{\mathrm{Gr(p,n)}}(X, S_i)^2$. This latter quantity is equal to $\sum_{i=1}^N \sum_{j=1}^p \theta_{ij}^2$ where θ_{ij} is the j-th principal angle between X and S_i. For $N > 2$, there is no closed-form solution for the mean subspace X. For this reason, the Riemannian mean is often approximated by the chordal mean [25], see Section 6. As it is well-known the sectional curvature of the Grassman manifold does not exceed 2, and the injectivity radius is $\pi/2$, we have guarantees that the Karcher mean exists and is uniquely defined in a geodesic ball of radius less than $\pi/(4\sqrt{2})$, see [20].

The Karcher mean of projectors in $\mathrm{Gr}(p,n)$ is a natural rank-preserving rotation-invariant mean that is well-defined on a subset of the boundary of the cone. We will use this mean subspace as a basis for the mean of N matrices of $\mathrm{S}^+(\mathrm{p},\mathrm{n})$.

4 A rank-preserving mean of an arbitrary number of matrices of $\mathrm{S}^+(\mathrm{p},\mathrm{n})$

The extension of the mean from projectors to arbitrary matrices of $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ is based on the decomposition

$$A = UR^2U^T,$$

of any matrix $A \in \mathrm{S}^+(\mathrm{p},\mathrm{n})$, with U an orthonormal matrix in $\mathrm{St}(\mathrm{p},\mathrm{n})$ and R^2 a positive definite matrix in $\mathrm{P_p}$. This matrix decomposition emphasizes the geometric interpretation of elements of $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ as *flat p-dimensional ellipsoids in $\mathbb{R}^n$*. The flat ellipsoid belongs to a p-dimensional subspace spanned by the columns of U, which form an orthonormal basis of the subspace, whereas the $p \times p$ positive definite matrix R^2 defines the shape of the ellipsoid in the low rank cone $\mathrm{P_p}$. The matrix decomposition $A_i = U_i R_i^2 U_i^T$ is defined up to an orthogonal transformation

$$U \mapsto UO, \qquad R^2 \mapsto O^T R^2 O\,,$$

with $O \in \mathrm{O(p)}$ since

$$A_i = U_i R_i^2 U_i^T = U_i O(O^T R_i^2 O) O^T U_i.$$

The orthogonal transformation does not affect the Grassman Riemannian mean but does affect, in general, the mean of the low-rank factors since $M(R_1^2, R_2^2) \neq M(R_1^2, O^T R_2^2 O)$ where M is the Ando mean. Principal difficulties for defining a proper geometric mean stem from this ambiguity.

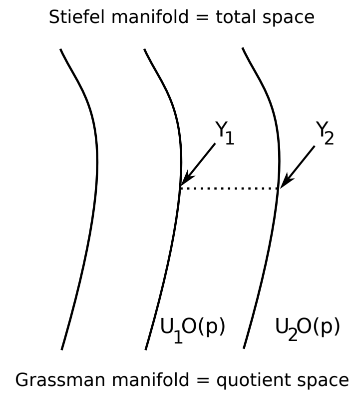

Figure 1: (Y_1, Y_2) are two bases of the subspaces spanned by the columns of U_1 and U_2 (the fibers) that minimize the distance in $\mathrm{St(p,n)}$. The dashed line represents the shortest path between those two fibers, thus its horizontal lift (i.e. its projection) in $\mathrm{Gr(p,n)}$ viewed as a quotient manifold, is a geodesic.

4.1 Mean of two matrices A_1 and A_2

Let $A_1 = U_1 R_1^2 U_1^T$ and $A_2 = U_2 R_2^2 U_2^T$ be elements of $\mathrm{S}^+(\mathrm{p,n})$. The representatives of the two matrices $(U_i, R_i^2), i = 1, 2$, are defined up to an orthogonal transformation

$$U_i \mapsto U_i O, \qquad R_i^2 \mapsto O^T R_i^2 O.$$

All the bases $U_i \mathrm{O(p)}$ correspond to the same p-dimensional subspace $U_i U_i^T$ (Figure 1). Note that, this representation of a p-dimensional subspace as the set of bases $U_i \mathrm{O}(p)$ is at the core of the definition of the Grassman manifold $\mathrm{Gr(p,n)}$ as a quotient manifold [22]

$$\mathrm{Gr(p,n)} \approx \mathrm{St(p,n)}/\mathrm{O(p)}.$$

The equivalence classes $U_i \mathrm{O(p)}$ are called the "fibers".

We will systematically assume that the principal angles between $\mathrm{span}(U_1)$ and $\mathrm{span}(U_2)$ are less than $\pi/2$, which is almost surely true if the subspaces $\mathrm{span}(U_1)$ and $\mathrm{span}(U_2)$ are picked randomly. In the case of two matrices, this is sufficient to ensure their Karcher mean in Gr(p,n) is unique. To remove the ambiguity in the definition of a mean of two PSD matrices, we propose to pick two particular representatives $Y_1 = U_1 Q_1$ and $Y_2 = U_2 Q_2$ in the fibers $U_1 \mathrm{O(p)}$ and $U_2 \mathrm{O(p)}$ by imposing that their distance in $\mathrm{St(p,n)}$ does not exceed the Grassman distance between the fibers they generate:

$$d_{\mathrm{St(p,n)}}(Y_1, Y_2) = d_{\mathrm{Gr(p,n)}}(\mathrm{span}(U_1), \mathrm{span}(U_2)), \tag{4}$$

Because the projection from $\mathrm{St(p,n)}$ to Gr(p,n) is a Riemannian submersion [18], and Riemannian submersions shorten the distances [26], this condition admits the equivalent formulation

$$(Q_1, Q_2) = \mathrm{argmin}_{(O_1, O_2) \in \mathrm{O(p)}} d_{\mathrm{St(p,n)}}(U_1 O_1, U_2 O_2). \tag{5}$$

which is illustrated by the picture of Figure 1: a geodesic in the Grasmman manifold admits the representation of a *horizontal* geodesic in $\mathrm{St(p,n)}$, that is, in the present case, a geodesic whose tangent vector points everywhere to a direction normal to the fiber.

The following proposition solves equation (5).

Proposition 2. *The compact SVD of $U_1^T U_2$ writes*

$$U_1^T U_2 = O_1 (\cos \Sigma) O_2^T . \tag{6}$$

where Σ is a diagonal matrix whose entries are the principal angles between the p dimensional subspaces [23]. If the pair (O_1, O_2) is defined via (6)*, it is a solution of* (5)*.*

Proof. We use a well-known result in the Grassman manifold: the shortest path between two fibers in $\mathrm{St(p,n)}$ concides with the geodesic path linking these two fibers in $\mathrm{Gr(p,n)}$, as the projection on the Grassman manifold is a Riemannian submersion, and thus shortens the distances (see [24, 26] for results on quotient manifolds). If two bases Y_1 and Y_2 of the fibers $U_1 O(p)$ and $U_2 O(p)$ are the endpoints of a geodesic in the Grassman manifold, they must minimize (5). It is thus sufficient to prove that $Y_1 = U_1 O_1$ and $Y_2 = U_2 O_2$ where O_1, O_2 are defined via (6) are the endpoints of a Grassman geodesic.

A geodesic in the Grassman manifold with Y_1 as starting point and Δ as tangent vector admits the general form [22]

$$\gamma(t) = Y_1 V \cos \Theta t V^T + U \sin \Theta t V^T , \tag{7}$$

where $U \Theta V^T = \Delta$ is the compact SVD of Δ. We thus propose to define the curve

$$Y(t) = Y_1 \cos \Sigma t + X \sin \Sigma t .$$

To define X we first assume all principal angles, i.e., all diagonal entries of Σ are strictly positive. Then we let $X = (Y_2 - Y_1 \cos \Sigma)(\sin \Sigma)^{-1}$. The curve $Y(t)$ is a geodesic, as it is of the form (7) with $\Delta = X\Sigma$ which is a tangent vector as $Y_1^T \Delta = 0$ (since $Y_1^T Y_2 = \cos \Sigma$), and $X\Sigma$ is a compact SVD as $X^T X = I$. This is because $X^T X = (Y_2^T - \cos \Sigma Y_1^T)(Y_2 - Y_1 \cos \Sigma)(\sin \Sigma)^{-2} = (I - (\cos \Sigma)^2)(\sin \Sigma)^{-2} = I$ where we used the fact that $Y_2^T Y_1 = Y_1^T Y_2 = \cos \Sigma$. Y_1 and Y_2 are its endpoints indeed as $Y_2 = Y(1)$. If there are null principal angles, it is clear that $Y(t)$ is a geodesic, where $X = (Y_2 - Y_1 \cos \Sigma)(\sin \Sigma)^\dagger$ along the directions corresponding to non-zero principal angles, and where X can be completed arbitrarily with orthonormal vectors along the directions corresponding to null principal angles. Indeed, along those directions Y_1 and Y_2, and thus $Y(t)$ coincide, and the value of X does not play any role in the definition of $Y(t)$.

$\square$

The following result allows to understand why the choice of the specific bases Y_1, Y_2 is relevant for defining a geometric mean, as explained in the end of this subsection. It proves the rotation of minimal energy (i.e. the closest to identity) mapping the subspace $\mathrm{span}(A_1)$ to $\mathrm{span}(A_2)$ maps Y_1 to Y_2.

Proposition 3. *Let $Y_1 = U_1 Q_1$ and $Y_2 = U_2 Q_2$ with (Q_1, Q_2) a solution of (6). Then the rotation $R \in \mathrm{SO(n)}$ that maps the basis Y_1 to the basis $Y_2 = RY_1$ is a rotation of minimal energy, that is, it minimizes $d_{\mathrm{SO(n)}}(R, I)$ among all rotation matrices that map Y_1 to the subspace* $\mathrm{span}(Y_2)$*.*

Proof. One can assume without loss of generality that $Y_1 = [e_1, \cdots, e_r]$ where $(e_1, \cdots, e_n)$ is the canonical basis of $\mathbb{R}^n$. Moreover, the search space can then be restricted to the rotations whose r first columns are of the form $Y_2 O$, whereas the $n - r$ remaining columns coincide with the identity matrix, as the rotation seeked must minimize the distance to identity. Any such rotation mapping Y_1 to $Y_2 O$ has its first columns equal to $Y_2 O$ and coincides with identity on the last $n - r$ columns. Thus we have for any such rotation $d_{\mathrm{St(n,n)}}(R, I) = d_{\mathrm{St(p,n)}}(Y_2 O, I)$. But as $\mathrm{SO(n)} = \mathrm{St(n,n)}$ and the metrics also coincide, we have $d_{\mathrm{SO(n)}}(R, I) = d_{\mathrm{St(n,n)}}(R, I)$. Thus the problem boils down to (5) and is solved taking $O = I$. $\square$

Having identified representatives as endpoints of a geodesic in Gr(p,n), their Riemannian mean in the Stiefel manifold (and in the Grassman manifold) is now easily written as the midpoint of the geodesic:

$$Y_1 \circ Y_2 = Y(t = \frac{1}{2}) = Y_1 \cos\frac{\Sigma}{2} + X \sin\frac{\Sigma}{2}\,. \tag{8}$$

Note that a weighted mean can be also computed using the geodesic parameterization:

$$Y_1 \circ Y_2 = Y(\alpha) = Y_1 \cos(\alpha\Sigma) + X \sin(\alpha\Sigma)\,, \tag{9}$$

where the weight given to Y_1 is $1-\alpha$ and the weight given to Y_2 is α.

Once Y_1 and Y_2 have been computed, R_1 and R_2 are given by the corresponding representatives

$$R_1^2 = Y_1^T A_1 Y_1, \qquad R_2^2 = Y_2^T A_2 Y_2. \tag{10}$$

The proposed mean of two matrices A_1, A_2 is then given by

$$A_1 \circ A_2 = W(R_1^2 \# R_2^2) W^T$$

where W is the Riemannian mean of Y_1 and Y_2 and $R_1^2 \# R_2^2$ is the Ando mean (2) of R_1^2 and R_2^2 in $\mathrm{P_p}$.

Proposition 3 provides a simple geometrical intuition underlying the definition of the mean: the mean of two flat ellipsoids A_1 and A_2 is defined in the mean subspace as the geometric mean of two full p-dimensional ellipsoids R_1^2 and R_2^2. There are several ways to rotate the ellipsoid A_1 into the subspace spanned by A_2. Different rotations will yield different respective positions of the two final ellipsoids. The choice is made univoque and sensible by selecting the minimal rotation. The rotated ellipsoid then merely writes $Y_2 R_1^2 Y_2^T$. Thus R_1^2 and R_2^2 define the shape of the ellipsoids expressed in the same basis Y_2. Figure 2 illustrates the proposed mean of two flat ellipsoids of $S^+(2,3)$.

4.2 Generalization to N matrices $A_1, \cdots, A_N \in \mathrm{S}^+(\mathrm{p},\mathrm{n})$

The construction presented in the previous section for two matrices is now extended to an arbitrary number of matrices. The main idea is to define a mean p-dimensional subspace and to bring all flat ellipsoids $A_1, \cdots, A_N$ to this mean subspace by a minimal rotation. In the common subspace, the problem boils down to compute the geometrical mean of N matrices in $\mathrm{P_p}$. The construction is achieved through the following three steps:

1. Let $A_i = U_i R_i^2 U_i^T$ for $1 \leq i \leq N$. Suppose that the subspaces spanned by the columns of the A_i's are enclosed in a geodesic ball of radius less than $\pi/(4\sqrt{2})$ in Gr(p,n). Then define the projector $W \in$ St(p,n) as a basis of their unique Karcher mean.

2. For each i, compute two bases Y_i and W_i of (respectively) $\mathrm{span}(U_i)$ and $\mathrm{span}(W)$ such that $d_{\mathrm{St(p,n)}}(Y_i, W_i) = d_{\mathrm{Gr(p,n)}}(\mathrm{span}(U_i), \mathrm{span}(W))$ i.e. solve problem (5). This is illustrated on Figure 3. Let $S_i^2 = Y_i^T A_i Y_i$. The ellipsoid A_i rotated to the mean subspace writes $W_i S_i^2 W_i^T$.

3. Let M denote the geometric mean alm or ls on $\mathrm{P_p}$. For each $1 \leq i \leq N$ let $T_i^2 = W_0^T W_i S_i^2 W_i^T W_0 \in \mathrm{P_p}$ where $W_0 \in \mathrm{St}(\mathrm{p},\mathrm{n})$ is a fixed basis of the mean subspace. The geometric mean of the matrices $A_1, \cdots, A_N$ is defined the following way:

$$A_1 \circ ... \circ A_N = W_0[M(T_1^2, \cdots, T_N^2)]W_0^T\,. \tag{11}$$

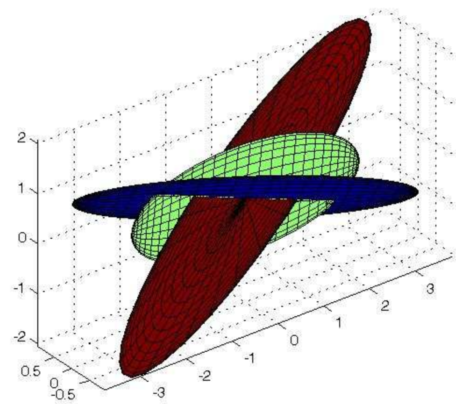

Figure 2: Proposed mean in $S^+(2,3)$. The proposed mean is such that both ellipsoids are brought to the mean subspace via a rotation of minimal energy, and then averaged. The resulting mean is a flat ellipsoid that lives in the mean subspace.

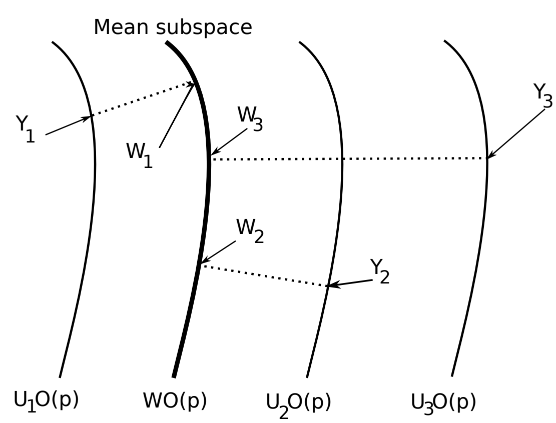

Figure 3: The bases $Y_1, \cdots, Y_N$ of the fibers and the bases $W_1, \cdots, W_N$ of the mean subspace fiber are such that (Y_i, W_i) are the endpoints of a geodesic in the Grassman manifold.

The approach can summarized as follows: in 1. a mean subspace is computed, in 2. the ellipsoids are rotated to this subspace, in 3. they are all expressed in a common basis W_0 so that their geometric mean can be computed in the small dimensional cone. An algorithmic implementation is proposed in Section 6.1.

5 Properties of the proposed mean of N matrices of $\mathrm{S}^+(\mathrm{p},\mathrm{n})$

Throughout this section

- M will systematically denote one of the two geometric means alm or ls on $\mathrm{P_p}$.
- it will be systematically assumed the subspaces spanned by the columns of $A_1, ...A_N \in \mathrm{S}^+(\mathrm{p},\mathrm{n})$ belong to a geodesic ball of radius less than $\pi/(4\sqrt{2})$ in Gr(p,n), so that the Karcher mean of these subspaces is well-defined and unique.
- with a slight abuse of notation, any projector UU^T where $U \in$St(p,n) will systematically be considered as an element of Gr(p,n), i.e. as a subspace.

5.1 Analytic properties

Proposition 4. *On the set of rank p projectors, the mean* (11) *coincides with the Grassman Riemannian mean. On the other hand, when the matrices in* $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ *all have the same range,* (11) *coincides with the geometric mean induced by the geometric mean M on the common range subspace of dimension p. More generally* (11) *coincides with M on the intersection of the ranges.*

Proof. The first two properties are obvious. The last one is linked to the special choice of a minimal energy rotation. Indeed, on the intersection of the ranges, the rotation of minimal energy is the identity. □

The next proposition proves that the proposed mean inherits the several properties of a geometric mean in the cone. For the reasons explained in the introduction of the paper, Properties (PP1) and (PP6-PP7), defined for the mean of three or more matrices in Subsection 2.2, must be adapted as follows: (PP1') if A, B, C commute and ABC has rank p, then $(A \circ B \circ C) = (A, B, C)^{1/3}$. (PP6') For $(\mu, P) \in \mathbb{R}_+^* \times \mathrm{O(n)}$ we have $(\mu P^T A \mu P) \circ (\mu P^T B \mu P) \circ (\mu P^T C \mu P) = \mu^3 P^T (A \circ B \circ C) P$. (PP7') $(A \circ B \circ C)^\dagger = A^\dagger \circ B^\dagger \circ C^\dagger$.

Proposition 5. *The mean* (11) *with* $M = \mathrm{alm}$ *is well-defined, and deserves the appellation "geometric" as it satisfies the properties (PP1'), (PP2-PP5), and (PP6'-PP7'). The result also holds with* $M = \mathrm{ls}$*, except that (PP4) can only be conjectured in this case.*

Proof. (PP1'): if $A_1 \circ \cdots \circ A_N$ has rank p, and $A_1, \cdots, A_N$ commute, then all A_i's span the same p-dimensional subspace. As a result, they all have the same range. On this common range, the mean has been proved to coincide with M, see Proposition 4. It thus inherits the (PP1) property.

M satisfies (PP2) and so does (11). To prove permutation invariance (PP3) it suffices to note that both Grassman mean and M are unvariant by permutation. To prove (PP4), suppose $A_i \leq A_i^0$ for each i. Then A_i's and A_i^0 have the same range and they all admit a factorization of the type $WR_i^2W^T$. (PP4) is then a mere consequence of the monotonicity of M. This property was proved for M=alm in [2] and it was conjectured for M=ls in [1]. Using the same arguments, one can prove continuity from above of the mean is a consequence

of continuity of M. (PP7') can be easily proved noting that for each i the pseudo-inverse writes $A_i^\dagger = U_i R_i^{-2} U_i^T$. Thus the calculation of the mean of the pseudo-inverse yields the inverse T_i^{-1}'s of the T_i's and (PP7') is the consequence of self-duality of M.

(PP6'): As for all $\mu > 0$ and i we have $\mu A_i = Y_i(\mu R_i^2)Y_i^T$ invariance with respect to scaling is a mere consequence of the invariance of M. Let $O \in \mathrm{O(n)}$. The mean subspace in Grassman of the rotated ranges (OA_iO^T)'s is the rotated mean subspace of the range of the A_i's. Proposition 2 says that $Y_i^T W_i = \cos\Sigma$ But for every i we have $(Y_i^T O^T)(OW_i) = Y_i^T W_i = \cos\Sigma$. Thus the matrices are transformed according to $W_i \mapsto OW_i$ and the $T_i's$ are unchanged. The mean of the rotated matrices is thus $OW_0\ G(T_1, ..., T_N)\ W_0^T O^T$. □

5.2 The proposed geometric mean as a Karcher mean in a special case

In the recent work [14], the authors proposed an extension of the affine-invariant metric of the cone to $\mathrm{S^+(p,n)}$. In this subsection, we explore the links between the Karcher mean in the sense of this metric and the proposed mean (11). We underline the fact that the proposed mean is *not* the Karcher mean in the cone. Yet, we prove that both means coincide in the meaningful case where all the matrices are rank p projectors.

The metric introduced in [14] is as follows. If $(U, R^2) \in \mathrm{St(p,n)} \times \mathrm{P_p}$ represents $A \in \mathrm{S^+(p,n)}$, the tangent vectors of $T_A\mathrm{S^+(p,n)}$ are represented by the infinitesimal variation (Δ, D), where

$$\begin{aligned} \Delta &= U_\perp B, \qquad B \in \mathbb{R}^{(n-p)\times p}, \\ D &= RD_0R, \end{aligned} \tag{12}$$

such that $U_\perp \in \mathrm{St(n, n-p)}$, $U^T U_\perp = 0$, and $D_0 \in \mathrm{Sym(p)} = T_I\mathrm{P_p}$. Vectors of the form (12) constitute a subset of tangent vectors to the total space $\mathrm{St(p,n)} \times \mathrm{P_p}$. This subset is called the horizontal space, and is defined such that each tangent of the horizontal space defines a unique tangent vector in the quotient $\mathrm{S^+(p,n)}$ (i.e. the horizontal space is transverse to the fibers). The chosen metric of $\mathrm{S^+(p,n)}$ needs only be defined on the horizontal space, and is merely the weighted sum of the infinitesimal distance in $\mathrm{Gr(p,n)}$ and in $\mathrm{P_p}$:

$$g_{k(U,R^2)}((\Delta_1, D_1), (\Delta_2, D_2)) = \mathrm{Tr}\left(\Delta_1^T \Delta_2\right) + k\ \mathrm{Tr}\left(R^{-1} D_1 R^{-2} D_2 R^{-1}\right),\ k > 0, \tag{13}$$

generalizing g^{P_n} in a natural way. The space $\mathrm{S^+(p,n)} \cong (\mathrm{St(p,n)} \times \mathrm{P_p})/\mathrm{O(p)}$ endowed with the metric (13) is a Riemannian manifold, and the metric is invariant to orthogonal transformations, scalings, and pseudo-inversion.

Proposition 6. *Consider N rank p projectors $U_1U_1^T, \cdots U_NU_N^T \in \mathrm{S^+(p,n)}$. Then the mean* (11) *is the Karcher mean of $U_1U_1^T, \cdots U_NU_N^T$ in the sense of metric* (13).

The proof of this propostion is based on two lemmas. Indeed, this result stems from the fact that (11) is of the form WW^T, where this latter projector is the Karcher mean of the N projectors in the sense of the Gr(p,n) natural metric. This means WW^T is the minimizer of the cost $G(VV^T) = \sum_i d^2_{\mathrm{Gr(p,n)}}(U_iU_i^T, VV^T)$. But the Karcher mean in $\mathrm{S^+(p,n)}$ is defined as the minimizer of the cost $F(X) = \sum_i d^2_{\mathrm{S^+(p,n)}}(U_iU_i^T, X)$. The first following lemma, proves that $F(X) \geq G(\mathrm{span}(X))$ for all $X \in \mathrm{S^+(p,n)}$. Thus for all X we have $F(X) \geq G(WW^T)$. But the second following lemma proves that $G(WW^T) = F(WW^T)$. As a result, $WW^T \in \mathrm{S^+(p,n)}$ minimizes F.

Lemma 1. *The distance between arbitrary A_1, A_2 in $\mathrm{S^+(p,n)}$ is lower bounded by the distance between their ranges in the Grassman manifold: $d_{\mathrm{S^+(p,n)}}(A_1, A_2) \geq d_{\mathrm{Gr(p,n)}}(\mathrm{span}(A_1), \mathrm{span}(A_2))$*

Proof. A horizontal curve $(U(t), R(t))$ has length $\int_0^1 \sqrt{g_{k(U(t),R(t))}(\dot{U}(t), \dot{R}(t))}dt$. For two matrices $A_1, A_2 \in \mathrm{S}^+(\mathrm{p},\mathrm{n})$, consider the horizontal lift $(U(t), R(t))$ of the geodesic linking A_1 and A_2 in $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ in the sense of metric (13). As the horizontal vector $(\dot{U}(t), \dot{R}(t))$ has a shorther norm in the tangent space than the horizontal vector $(\dot{U}(t), 0)$, we have $d_{\mathrm{S}^+(\mathrm{p},\mathrm{n})}(A_1, A_2) \geq d_{\mathrm{St}(\mathrm{p},\mathrm{n})}(U(0), U(1))$. Besides, $U(t)$ defines a curve in $\mathrm{St}(\mathrm{p},\mathrm{n})$ linking $\mathrm{span}(A_1)$ and $\mathrm{span}(A_2)$. As the projection from the Stiefel manifold to the Grassman manifold viewed as a quotient space $\mathrm{Gr}(\mathrm{p},\mathrm{n}) \simeq \mathrm{St}(\mathrm{p},\mathrm{n})/\mathrm{O}(\mathrm{p})$ is a Riemannian submersion, it shortens the distances[26], i.e. $d_{\mathrm{Gr}(\mathrm{p},\mathrm{n})}(\mathrm{span}(A_1), \mathrm{span}(A_2)) \leq d_{\mathrm{St}(\mathrm{p},\mathrm{n})}(U(0), U(1))$. This proves the result. □

Lemma 2. *In the particular case where A_1, A_2 in $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ are two projectors, the geodesic joining them in $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ coincides with the geodesic joining their ranges in $\mathrm{Gr}(\mathrm{p},\mathrm{n})$. It implies $d_{\mathrm{S}^+(\mathrm{p},\mathrm{n})}(A_1, A_2) = d_{\mathrm{Gr}(\mathrm{p},\mathrm{n})}(\mathrm{span}(A_1), \mathrm{span}(A_2))$.*

Proof. One can find a horizontal curve in $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ whose length is $d_{\mathrm{Gr}(\mathrm{p},\mathrm{n})}(A_1, A_2)$, by choosing representatives in $\mathrm{St}(\mathrm{p},\mathrm{n})$ as in Proposition 2. It is thus a geodesic in Grassman, but also in $\mathrm{S}^+(\mathrm{p},\mathrm{n})$ because of Lemma 1. □

Beyond the particular case of projectors, it must be emphasized that the mean (11) is not the Karcher mean in the sense of metric (13). This is because a horizontal curve $(U(t), R(t))$ that is made of a geodesic $U(t)$ in Grasmman and of a geodesic $R(t)$ in the cone does not define a geodesic in $\mathrm{St}(\mathrm{p},\mathrm{n})$. For instance, it is possible to construct a geodesic joining matrices having the same range, and such that the mid-point does not have the same range (see [14], Proposition 1). This counter-example shows Proposition 4, although very natural, is not satisfied by the Karcher mean, as the mean of matrices having the same range does not boil down to their geometric mean within this range. Even if the metric seems natural, and has been successfully used in several applications (e.g., [27, 28]), the resulting Karcher mean lacks elementary expectable properties that the mean (11) possesses.

6 Application to filtering

6.1 Algorithmic implementation and computational cost

Here we recap the basic steps for an implementation of the mean. The calculation of the mean has a numerical complexity of order $O(np^2)$. This cost is linear with respect to n, a very desirable feature in large-scale applications where $n \gg p$.

1. For $1 \leq i \leq N$ let U_i be any orthonormal basis of the span of A_i.

2. Let W be an orthonormal basis of the subspace that is the Karcher mean in the Grassman manifold between the associated subspaces. Instead of minimizing $\sum_{i=1}^{N}\sum_{j=1}^{p}\theta_{ij}^2$, an interesting alternative is to minimize $\sum_{i=1}^{N}\sum_{j=1}^{p}(\sin\theta_{ij})^2$, which corresponds to approximate the angular distance by a chordal distance. Both definitions are asymptotically equivalent for small principal angles. In this case, the solution was shown in [25] to be the p-dimensional dominant subspace of the centroid $\sum_{i=1}^{N} U_i U_i^T$, which can easily be found by truncated SVD.

3. For $1 \leq i \leq N$

   - The SVD of $U_i^T W$ yields two orthogonal matrices O_i, O_i^W such that $O_i^T U_i^T W O_i^W$ is a diagonal matrix.

- Let $Y_i = U_i O_i$ and $W_i = W O_i^W$. Let $S_i^2 = Y_i^T A_i Y_i$. Let $T_i^2 = W^T W_i S_i^2 W_i^T W$.

4. Compute the geometric mean in the low-rank cone $M(T_1^2, \cdots, T_k^2)$ using methods in the literature [2, 21].

5. The geometric mean is: $W\ M(T_1^2, \cdots, T_k^2)\ W^T$.

6.2 Geometric means and filtering applications

Filtering on $S^+(n, n)$ with the metric (13) (which is the GL(n)-invariant metric of the cone $\mathrm{P_n}$) was studied extensively for diffusion tensor images (DTI) filtering in [6, 7, 10]. It was also applied to signal processing in [13], and also seems to be promising in radar processing [29]. One of the main benefits of this metric is its invariance with respect to scalings which makes it very robust to outliers, i.e. large noise, as the effect of a large eigenvalue is mitigated by the geometric mean. This very property, which is desirable for means in the interior of the cone, yields a great lack of robustness to (even small) noises as soon as some matrices are rank-deficient. The mean (11) inherits the invariance to scalings property, which yields robustness to outliers, without being subject to the same problems in case of rank deficiency, as illustrated by the following proposition.

Proposition 7. *Let $A \in \mathrm{S}^+(\mathrm{p}, \mathrm{n})$, and R_ϵ be a rotation of magnitude ϵ. If $\mathrm{span}(R_\epsilon A) \cap \mathrm{span}(A) = \emptyset$, which can be the case with arbitrary small $\epsilon > 0$ as soon as $p < n/2$, the Ando mean of A and $R_\epsilon A$ is the null matrix according to Propostion 1. On the other hand, $A \circ R_\epsilon A \to A$ when $\epsilon \to 0$.*

This proposition shows that the Ando mean of a stream of noisy measurements $R_\epsilon A$ of the matrix A, is generally the null matrix, even with arbitrarily small noises, whereas it should be close to A. On the other hand, it is indeed the case when the rank-preserving metric proposed in this paper is used. This appears to be a fundamental feature in filtering applications.

6.3 A toy example: filtering a constant rank p positive semi-definite matrix

In continuous-time, a first-order filter meant to filter a constant noisy signal $y(t)$ writes

$$\tau \frac{d}{dt} x = -x + y\,.$$

In discrete-time, using a semi-implicit numerical scheme, it becomes

$$x_{i+1} = \frac{y_i dt + \tau x_i}{dt + \tau}\,, \tag{14}$$

which is a weighted mean between the measured signal y_i and the filtered signal x_i. Thus a weighted version of the mean proposed in this paper can be used to do filtering. Indeed if the usual matrix arithmetic mean is used, the rank is not preserved.

For the sake of illustration, we apply this idea to a weighted version of the mean proposed in this paper applied to the filtering of a constant matrix zz^T of $\mathrm{S}^+(1, 2)$, where $z \in \mathbb{R}^2$ is a constant vector. Let $Y(t) = (z + \nu(t))(z + \nu(t))^T \in \mathrm{S}^+(1, 2)$ where $\nu(t) \in \mathbb{R}^2$ is a Gaussian white noise of magnitude 50% of the signal. The filtering algorithm based on a weighted

mean corresponding to the metric (13) writes:

$$u_{i+1} = (\cos(\frac{\tau\theta_i + \theta_{Y_i} dt}{dt+\tau}), \sin(\frac{\tau\theta_i + \theta_{Y_i} dt}{dt+\tau}))\,, \tag{15}$$

$$r_{i+1}^2 = \exp(\frac{\tau \log r_i^2 + dt \log s_i^2}{dt+\tau})\,, \tag{16}$$

where the current estimated matrix is $r_i^2 u_i u_i^T$ with $u_i = (\cos\theta_i, \sin\theta_i)$, and $Y_i = s_i^2 v_i v_i^T$ where $v_i = (\cos\theta_{Y_i}, \sin\theta_{Y_i})$. The Grassman mean was computed using (9). The result is represented on figure 4.

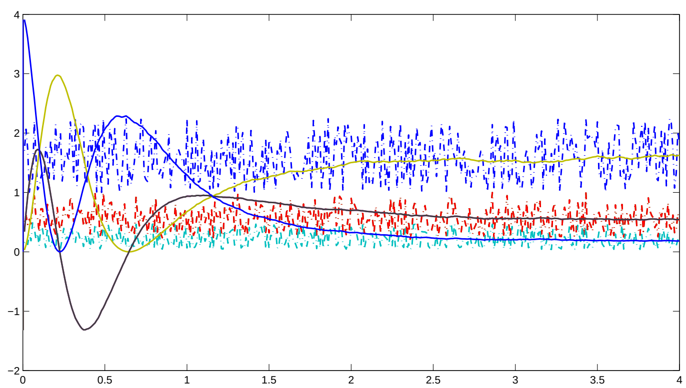

Figure 4: Filtering on $S^+(1,2)$: plot of the 3 coefficients of the measured matrix (dashed line) and the filtered matrix (plain line) with a 50% measurement noise, and $\tau = 50dt$. The filter allows to denoise the measured rank-1 symmetric matrix.

As mentioned earlier, the geometric mean in the cone has proved useful in several filtering applications due to its remarkable robustness to outliers. Here, it is easy to see for instance in the $S^+(1,2)$ case looking at update (16), that an outlier having a very large amplitude will be crushed thanks to the logarithm function. Such a property is inherited by the metric (13) for filtering on $\mathrm{S}^+(\mathrm{p},\mathrm{n})$. See Figure 5 for an experimental illustration of the property.

Acknowledgements

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

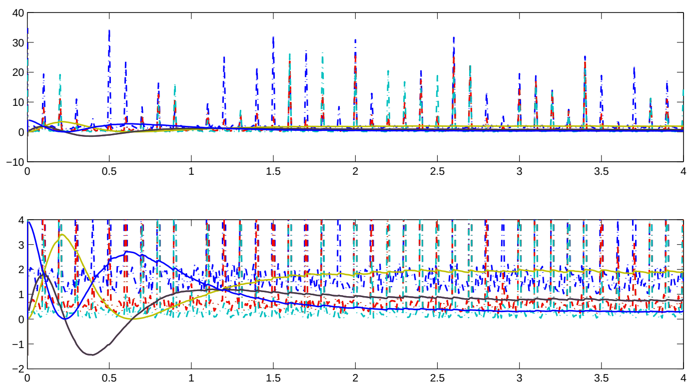

Figure 5: Above : same setting as Figure 4 but an outlier is inserted at each 10 steps. Below : zoom on the filtered signals (to be compared to Figure 4). The filter shows good robustness properties to outliers.

References

[1] R. Bhatia, Positive definite matrices, Princeton University Press, Princeton, NJ, 2006.

[2] T. Ando, C.K. Li, R. Mathias, Geometric means, Linear Algebra Appl. 385 (2004) 305–334.

[3] D. Petz, R. Temesi, Means of positive numbers and matrices, SIAM J. matrix anal. appl. 27 (2005) 712–720.

[4] R. Bhatia, J. Holbrook, Riemannian geometry and matrix geometric means, Linear Algebra Appl. 413 (2006) 594–618.

[5] J. Faraut, A. Koranyi, Analysis on Symmetric Cones, Oxford University Press, New York, 1994.

[6] X. Pennec, P. Fillard, N. Ayache, A Riemannian framework for tensor computing, International Journal of Computer Vision. 66 (2006) 41–66.

[7] P. D. Fletcher, S. Joshi, Riemannian geometry for the statistical analysis of diffusion tensor data, Signal Processing. 87 (2007) 250–262.

[8] M. Zerai, M. Moakher, The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of diffusion tensor MRI data, J. Math. Imaging Vision, submitted.

[9] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (2005) 735–747.

[10] V. Arsigny, P. Fillard, X. Pennec, N. Ayache. Geometric mean in a novel vector space structure on symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 29 (2007) 328–347.

[11] J. Burbea, C. R. Rao, Entropy differential metric, distance and divergence measures in probability spaces: A unified approach, J. Multivariate Anal. 12 (1982) 575–596.

[12] L. T. Skovgaard, A Riemannian geometry of the multivariate normal model, Scand. J. Statist. 11 (1984) 211–223.

[13] S.T. Smith, Covariance, subspace, and intrinsic Cramer-Rao bounds, IEEE-Transactions on Signal Processing. 53(2005) 1610–1629.

[14] S. Bonnabel, R. Sepulchre, Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank, SIAM J. Matrix Anal. Appl. 31(2009) 1055–1070.

[15] S. Bonnabel, R. Sepulchre, Rank-preserving geometric means of positive semi-definite matrices, International Symposium on Mathematical Theory of Networks and Systems (MTNS). (2010) 209–2014.

[16] T. Ando, Topics on operator inequalities, In Division of Applied Mathematics, Hokkaido University, Sapporo, Japan, 1978.

[17] W. Pusz, S. L. Woronowicz, Functional calculus for sesquilinear forms and the purification map, Rep. Mathematical Phys.8 (1975) 159–170.

[18] P.-A. Absil, R. Mahony, R. Sepulchre. Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ, 2007.

[19] H. Karcher, Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math., 30 (1977), pp. 509–541.

[20] B. Afsari, Riemannian Lp center of mass : existence, uniqueness, and convexity. Proceedings of the American Mathematical Society, S 0002-9939(2010)10541-5, Article electronically published on August 27, 2010.

[21] M. Arnaudon, C. Dombry, A. Phan and Le Yang, Stochastic Algorithms for computing means of probability measures. Preprint, Arxiv, 2011

[22] A. Edelman, T. A. Arias, S. T. Smith, The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. 20(1998), 303–353.

[23] G.H. Golub, C. Van Loan, Matrix Computations, John Hopkins University Press, Baltimore, 1983.

[24] B. O'Neill, Semi-Riemaniann geometry, Pure and applied mathematics. Academic Press Inc., New York, 1983.

[25] A. Sarlette, R. Sepulchre, Consensus optimization on manifolds, SIAM J. Control and Optimization, 48 (2009) 56-76.

[26] S. Gallot, D. Hulin, J. Lafontaine, Riemannian geometry. Springer, Berlin, 2004.

[27] G. Meyer, S. Bonnabel, R. Sepulchre, Regression on fixed-rank positive semidefinite matrices: a Riemannian approach. Journal of Machine Learning Reasearch (JMLR). 12(Feb):593-625, 2011.

[28] S. Bonnabel, R. Sepulchre, The geometry of low-rank Kalman filters. To appear. Matrix Information Geometry. Springer Verlag, 2012.

[29] F. Barbaresco, Innovative tools for radar signal processing based on Cartan's geometry of symmetric positive-definite matrices and information geometry. IEEE Radar Conference, 2008,

[30] P. Perona, J. Malik. Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence. 12 (1990) 629–639.